

METHODS AND APPARATUS FOR INTELLIGENT CRAWLING
ON THE WORLD WIDE WEB

Abstract of the Disclosure

Methods and apparatus for performing intelligent crawling are provided. Particularly, the intelligent crawling techniques of the invention provide a crawler mechanism which is capable of learning as it crawls in order to focus the search for documents on the information network being explored, e.g., world wide web. This crawler mechanism stores information about the crawled documents as it retrieves the documents, and then uses the information to further focus its search appropriately. The inventive techniques result in the crawling of a small percentage of the documents on the world wide web.

1500-150.APP